Numu CC – NC Separation using Artificial Neural Networks

N. Saoulidou, Fermilab G.Tzanakos & M. Zois, Univ. of Athens

MINOS Collaboration Meeting, March 26, 2004

From September 2003

Future Work Plan

- **Expand** the set of input variables
- Use **tracking/shower** information
- Study ANN performance versus E_{vis} instead of E_{ν} (which is unknown!)
- Learn how to **interpret** the classification results in order to obtain **"signal"** and **"background"** at a given **CL**.
- Attempt to apply for the estimation of oscillation limits and parameters.

Outline

- Goal of the analysis
- Method
  - Ann Basics
- Results
- Conclusions – On going work

Goal – Physics Motivation

- **Goal :**
  - Classify neutrino events into numu CC – nue CC – NC
- **Physics Motivation :**
  - Numu disappearance
  - Nue appearance
  - Other senarios….
- **Needs to be performed for both Near & Far data**

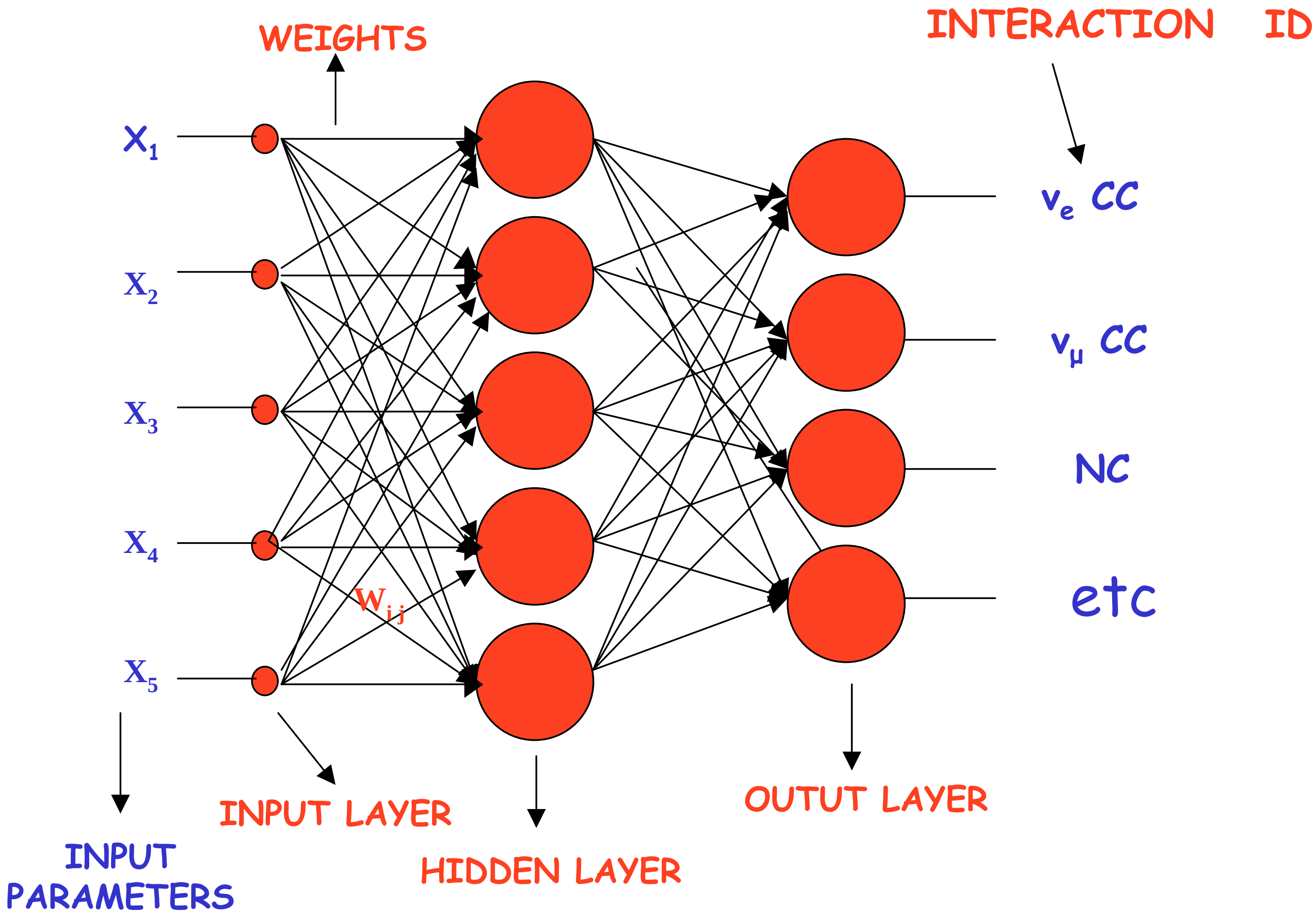
Continuing from previous work

- Expand the set of input variables
- Use tracking/shower information
- Study ANN performance versus E_{vis} instead of E_{ν} (which is unknown!)
- Learn how to interpret the classification results in order to obtain "signal" and "background" at a given CL.
- Attempt to apply for the estimation of oscillation limits and parameters.
- Study the significance of correct a priori probabilities to the ANN results.

Method - Strategy

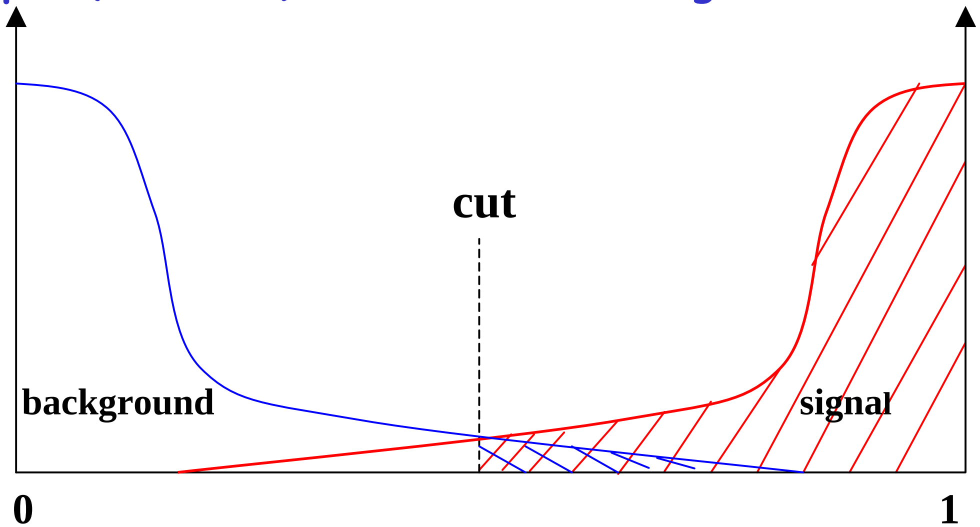
- Generated & reconstructed 10.000 mono-energetic neutrino events (numu CC & NC) in the FD for:
  - 0.5 GeV 1 GeV 2GeV 3 GeV 5 GeV 10 GeV
- Used Nathaniel's DetSim and frozen release R1.5
- Divided the events (after preprocessing them) into visible energy "bins":
  - 0-200 ADC counts 200-400 ADC counts and 400-… ADC counts
- Constructed a separate ANN (MLPfit) for each different Evis. range

ANN Schematic



ANN Parameters

Network output (selection) function for “background ”and “signal” events



S = Total # Signal events

B = Total # Background events

S_C = Signal events above Cut

B_C = Background events above Cut

$$\text{efficiency} = \frac{S_C}{S}$$

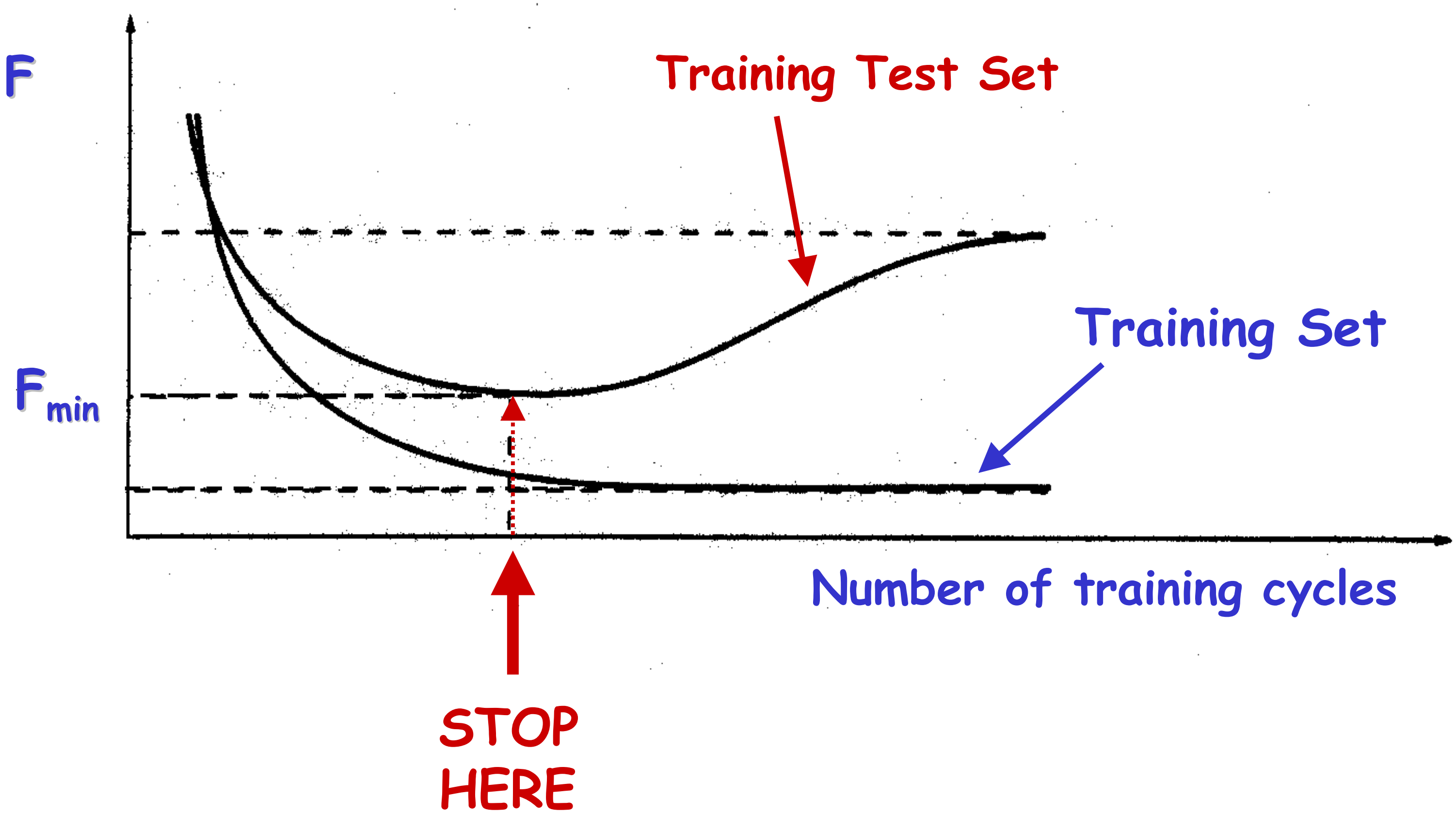
$$\text{purity} = \frac{S_C}{S_C + B_C}$$

$$\text{contamination} = \frac{B_C}{B}$$

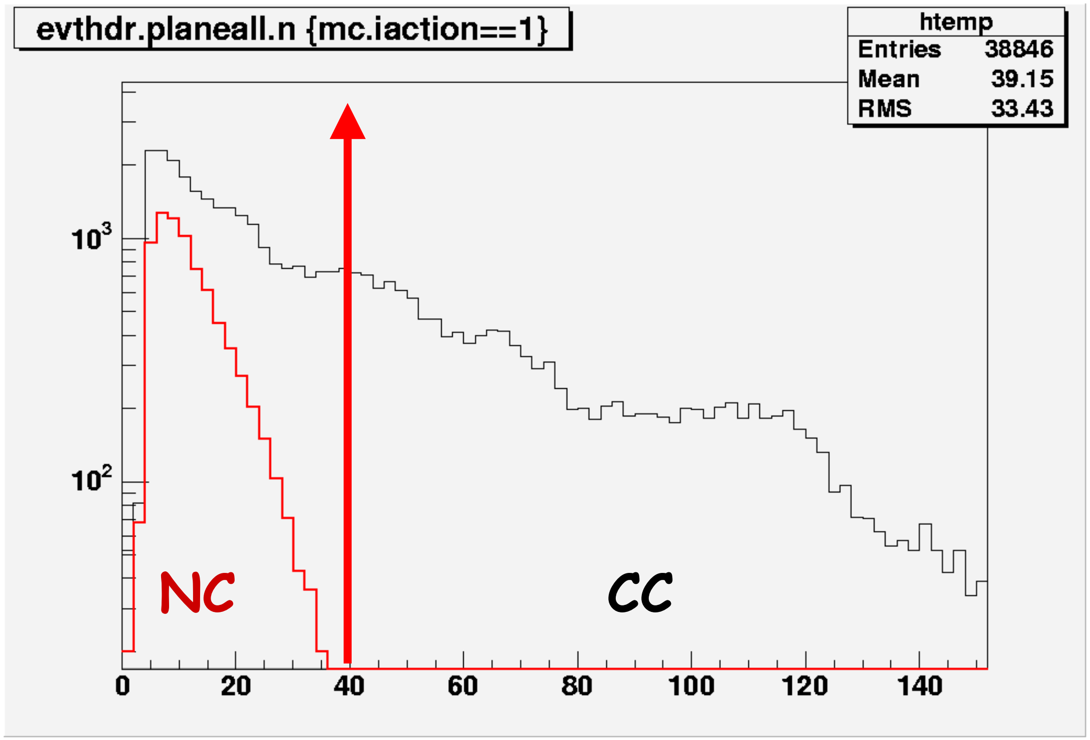
Overtraining (Early stopping method)

F is the "cost function"

When an ANN get overstrained is looses its generalization ability and learns ONLY the specific training examples



Preprocessing of the Events

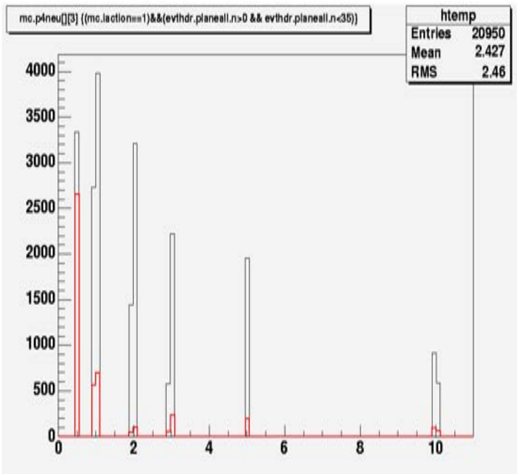


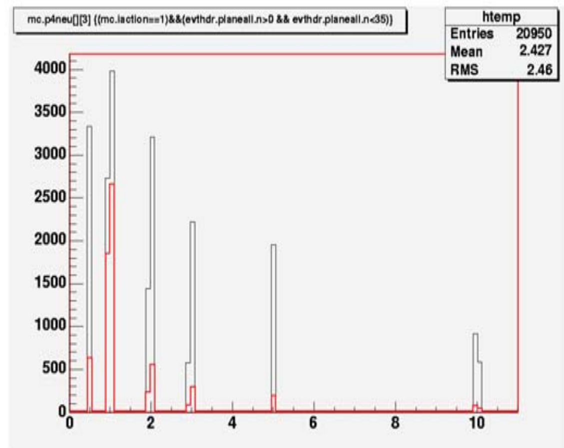
- Placing an initial simple cut at the number of planes we select 46 % of the Signal & ~ 0% of the Background

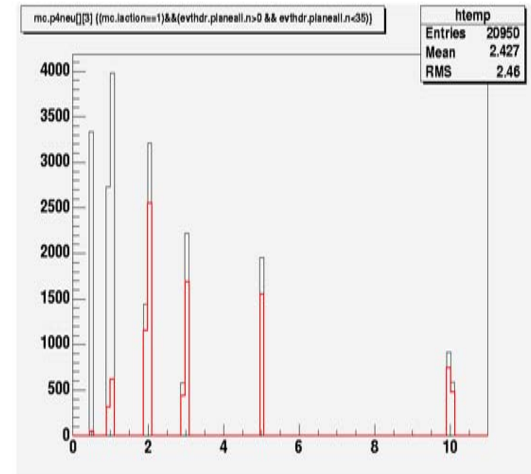
Relation between E visible & E true

RED : Selected in each Evis Range BLACK: ALL

Numu CC

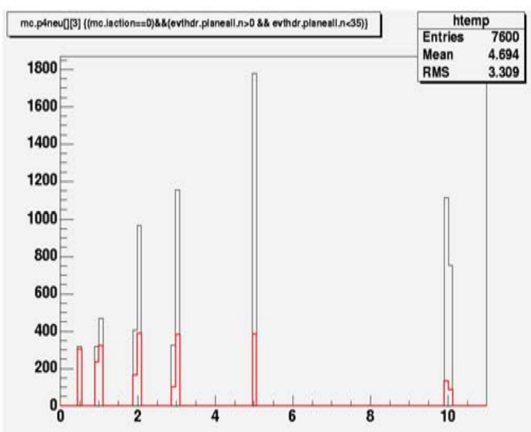


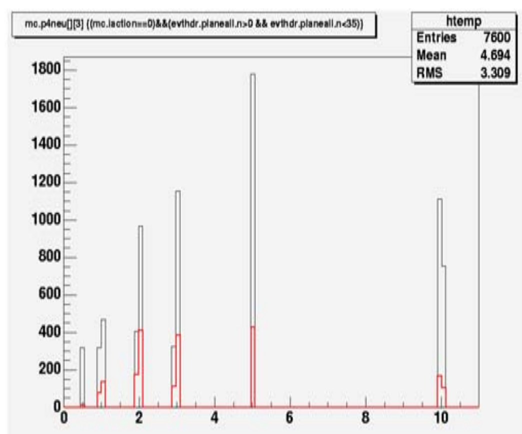


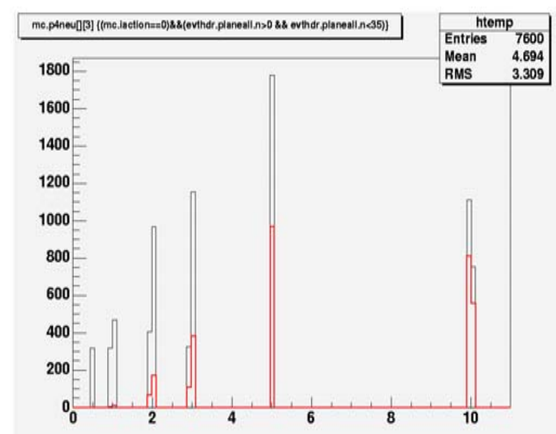


<200ADC counts 200 – 400 ADC counts > 400 ADC counts

NC







E (GeV)

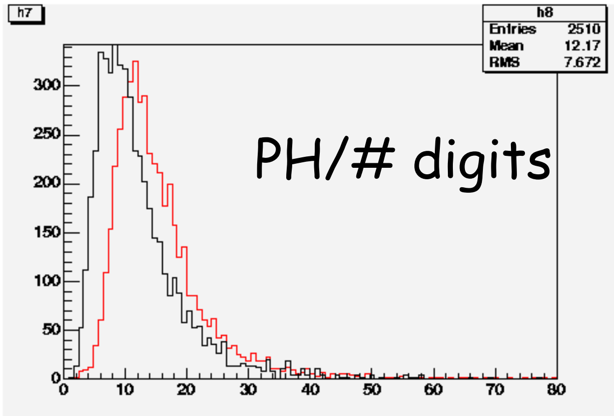
ANN architecture & Input variables

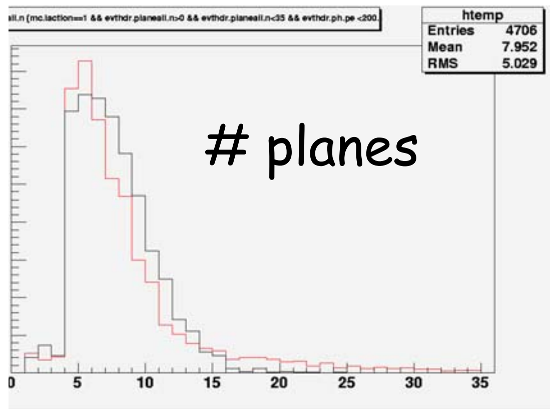
ANN Architecture
13 inputs
1 hidden layer with 8 neurons
1 output

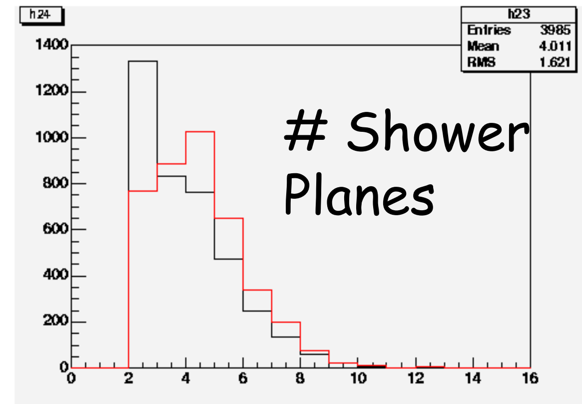
Input Variables :
Pulse height per plane
Pulse height per strip
Pulse height per digit
Number of planes
Number of strips
Number of tracks
Number of showers
Shower energy
Shower number of planes
Percentage of shower energy to total event energy
Shower energy per plane
Shower energy per strip
Transverse Plane asymmetry

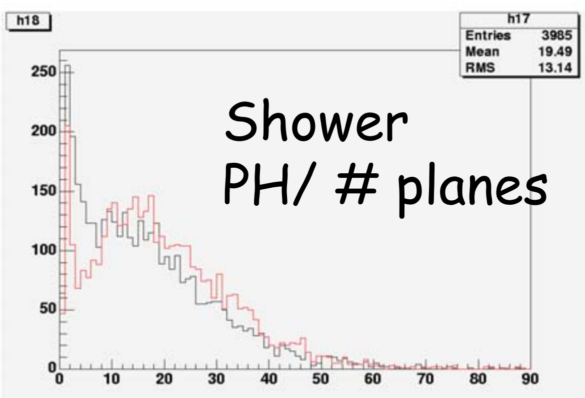
Numu CC – NC , 0 – 200 ADC counts, variables

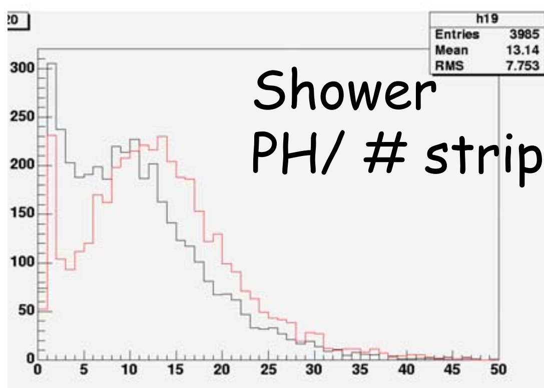
Red : Numu CC Black : NC

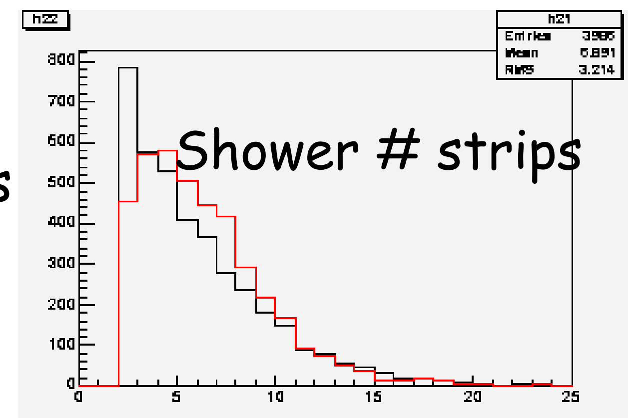










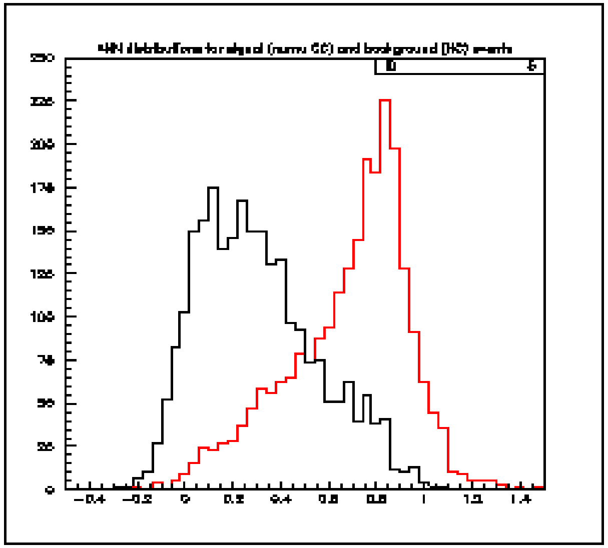


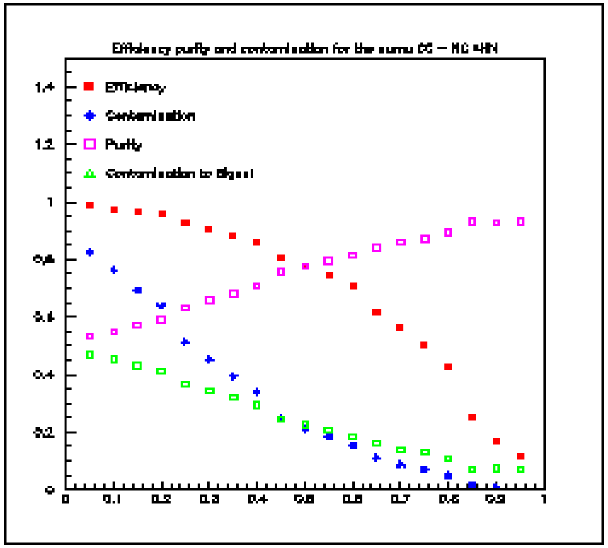
- Most of the variables show very slight differences.

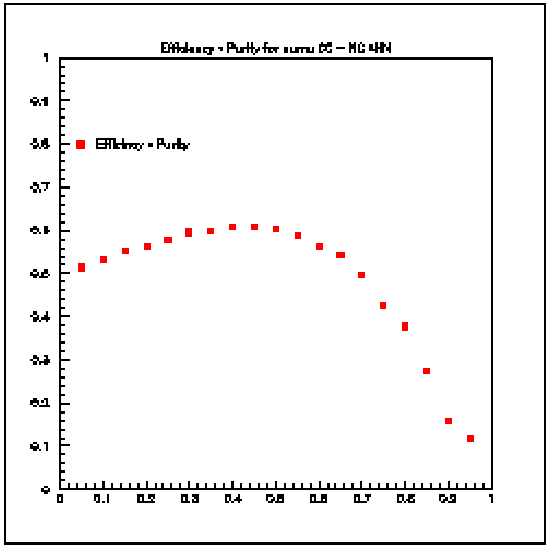
Numu CC - NC , 0 - 200 ADC counts, Results

Efficiency Purity Contamination

Efficiency x Purity







Event Probability

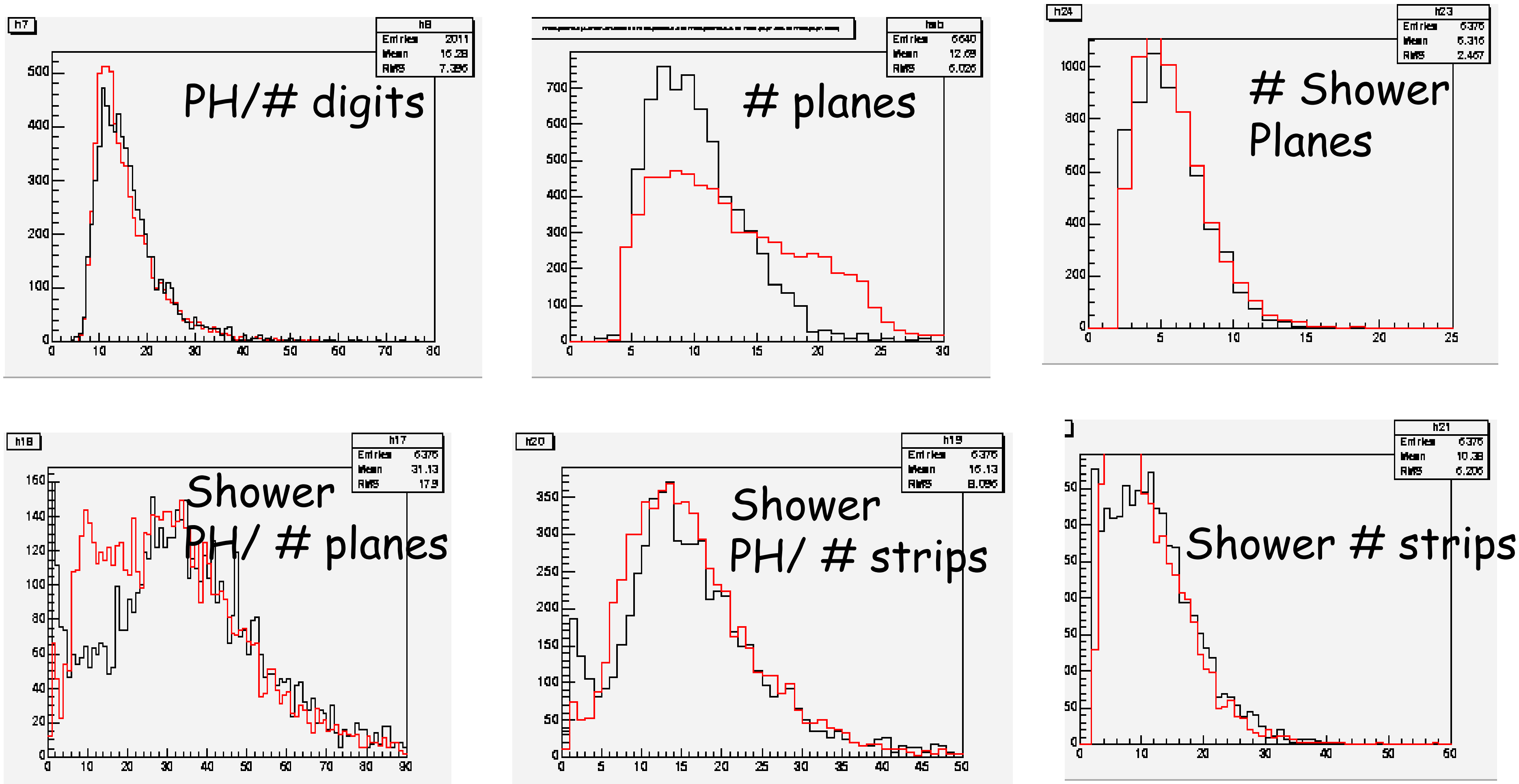
Cut

Cut

- The ANN discriminating power is quite high :
  - cut @ 0.5 Efficiency 77 % Purity 78 % (Numu CC)

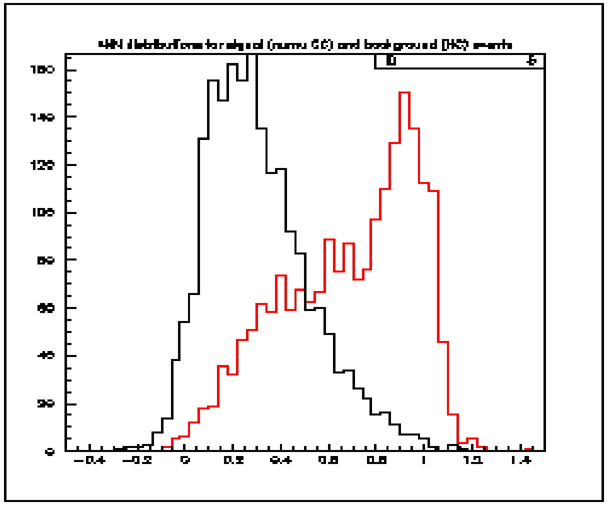
Numu CC – NC , 200 – 400 ADC counts, variables

Red : Numu CC Black : NC



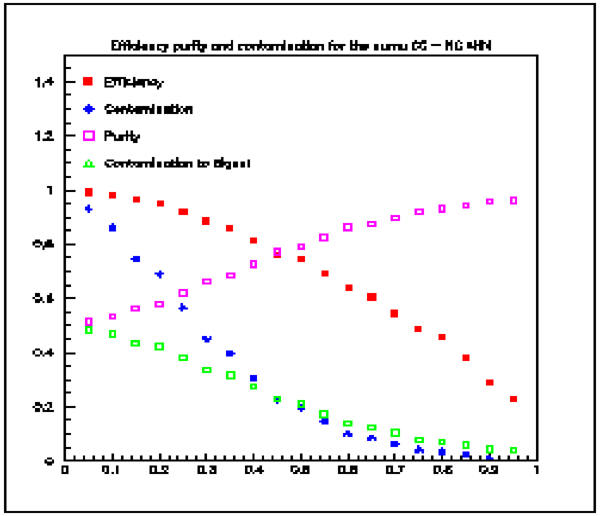
- Most of the variables show very small differences.

Numu CC - NC , 200 - 400 ADC counts, Results



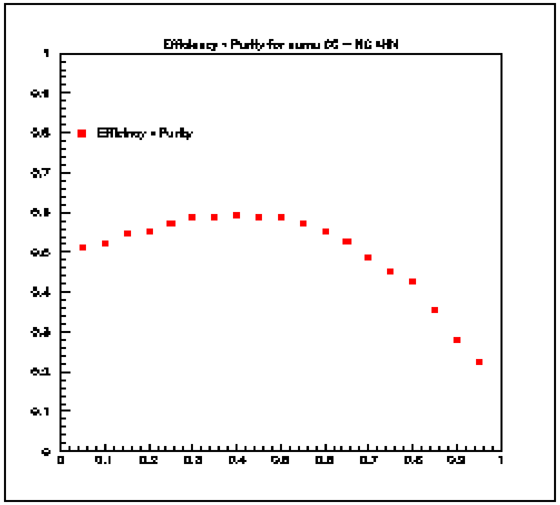
Event Probability

Efficiency Purity Contamination



Cut

Efficiency × Purity

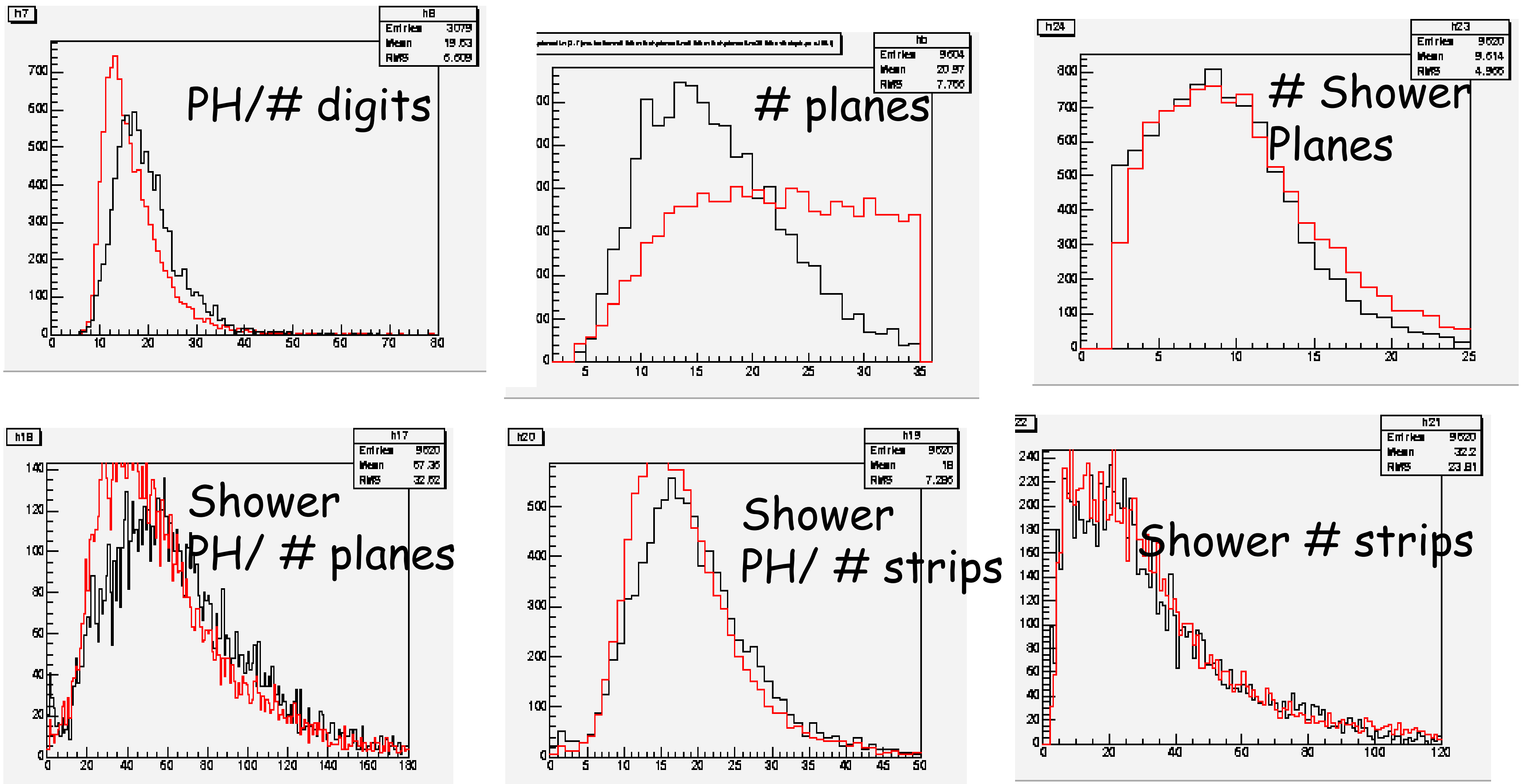


Cut

- **The ANN discriminating power is quite high :**
  - **cut @ 0.5 Efficiency 75 % Purity 79 % Numu CC**

Numu CC – NC , > 400 ADC counts, variables

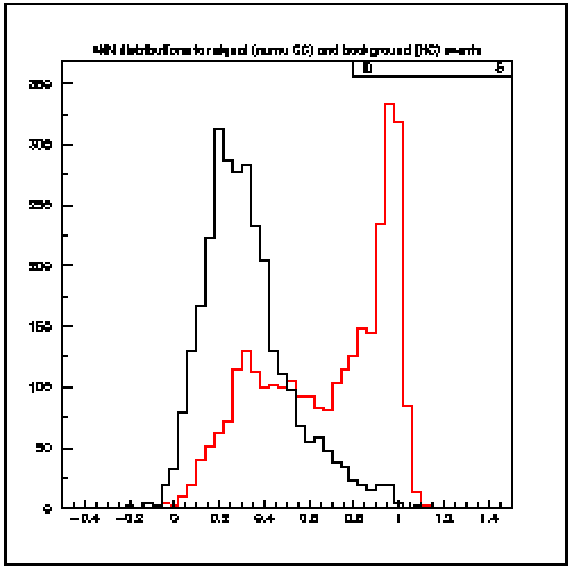
Red : Numu CC   Black : NC

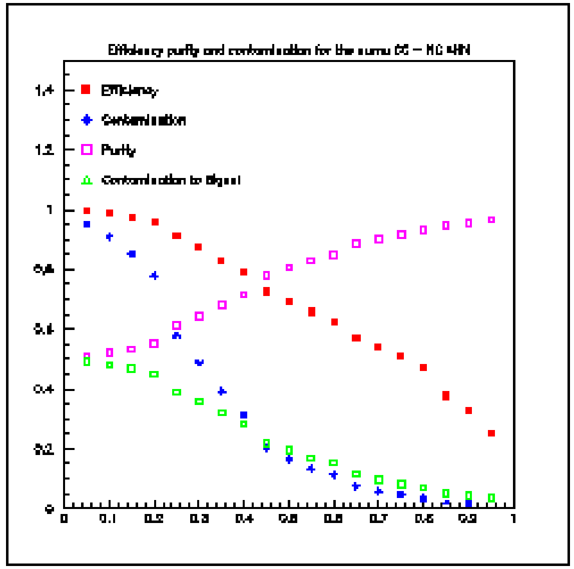


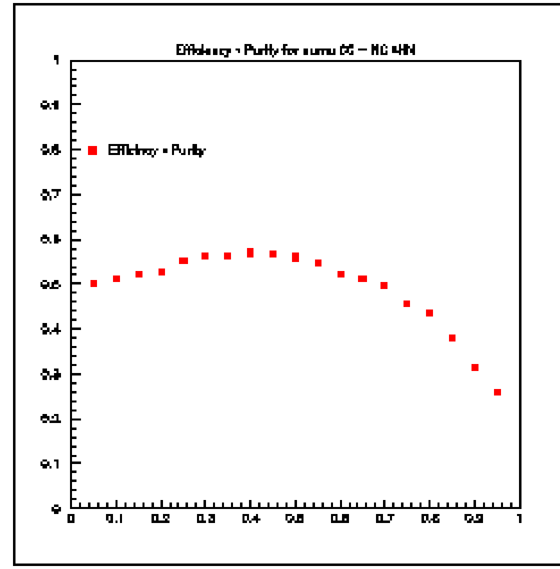
- Most of the variables show very small differences.

Numu CC - NC , > 400 ADC counts, Results

Efficiency Purity Contamination   Efficiency x Purity





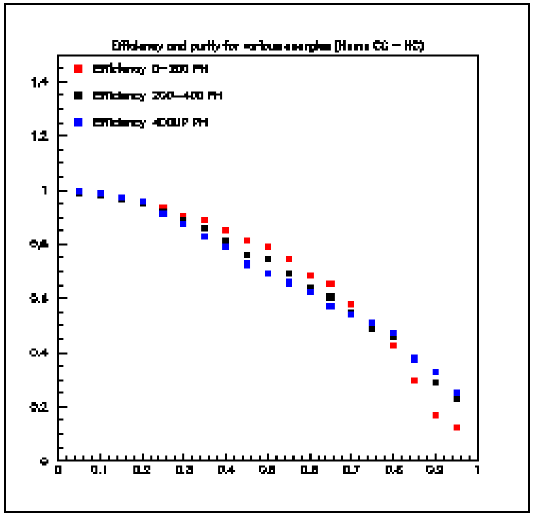


Event Probability   Cut   Cut

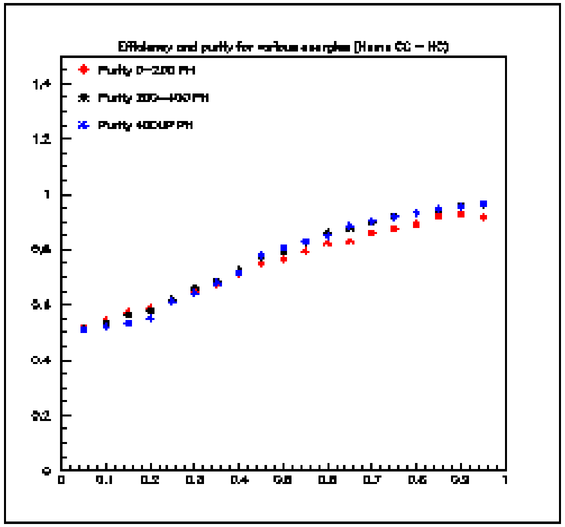
- **The ANN discriminating power is quite high :**
  - **cut @ 0.5 Efficiency 70 %   Purity 80 % Numu CC**

Comparisons 0-200 ,200-400 & 400 > ANNs

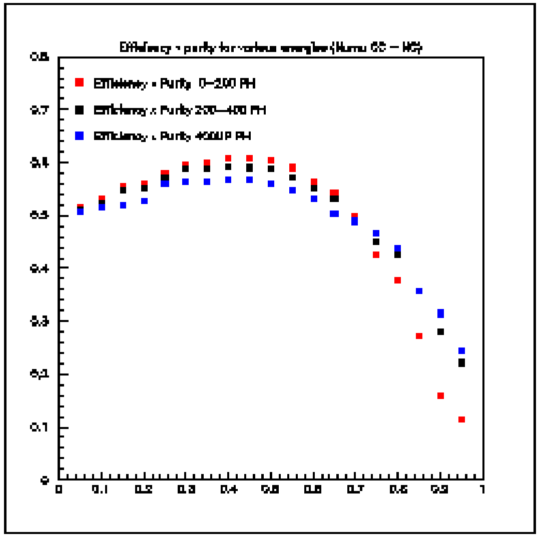
Efficiency



Purity



Eff x Pur



- The results for the three different visible energy ranges are very similar.
- The efficiency seems to be slightly decreasing while the purity is increasing as Evis. increases.

ANN ε and p for N_s & N_B calculation

- Having determined efficiency and purity for each visible energy range we can extract the “true” number of signal and background events for this particular range:

N = # of events in this range.

Ns = # of true signal events.

Ns-sel = # of “signal-like” selected events

ε = Ns-selected/Ns

p = Ns-selected/N

Ns-sel=p*N=>Ns=Ns-sel*ε =>Ns=pN/ε

ANN probability (review)

- **Bayesian a posteriori probability :**

$$P(S/x) = \frac{P(x/S) * P(S)}{(P(S) * P(x/S) + P(B) * P(x/B)}$$

$P(S) = \text{a priori} \quad \text{signal probabilty}$ $\qquad P(x/S) = \text{Signal probabilty density function}$

$P(B) = \text{a priori} \quad \text{background probabilty}$ $\qquad P(x/B) = \text{Background probabilty density function}$

- **ANN output : P(S/x)**
- **ANN training examples : P(x/S) & P(x/B)**
- **ANN number of Signal Training Examples P(S)**
- **ANN number of Background Training Examples P(B)**

The MLP (ann) analysis and the Maximum Likelihood Method (Bayes Classifier) are equivalent.

(c_{11} c_{22} = cost for making the correct decision &

c_{12} c_{21} = cost for making the wrong decision)

$$\Lambda(x) = \frac{P(x/S)}{P(x/B)} \;\&\; \xi = \frac{P(B)(c_{12} - c_{11})}{P(S)(c_{21} - c_{22})}$$

$$\text{if} \quad c_{11} = c_{22} = 0 \;\&\; c_{12} = c_{21} \Rightarrow$$

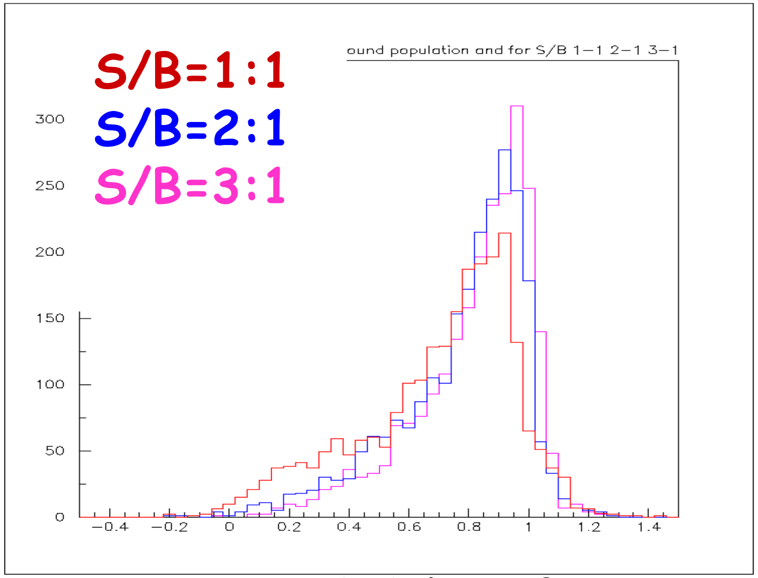
$$\Lambda(x) > \xi \Leftrightarrow \frac{P(x/S)}{P(x/B)} > \frac{P(B)}{P(S)} \Leftrightarrow P(x/S) * P(S) > P(x/B) * P(B) \Leftrightarrow$$

$$\Leftrightarrow \frac{P(x/S) * P(S)}{P(x)} > \frac{P(x/B) * P(B)}{P(x)} \Leftrightarrow P(S/x) > P(B/x) \Leftrightarrow$$

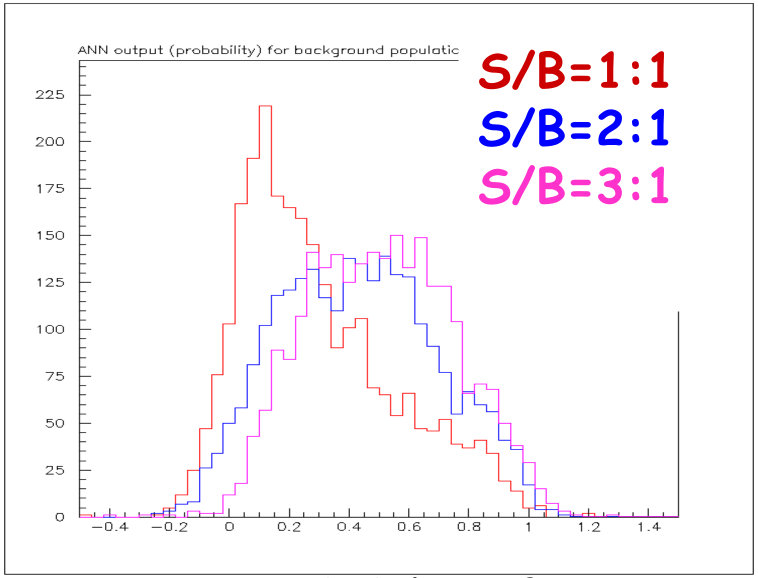
$$\Leftrightarrow P(S/x) > (1 - P(S/x)) \Leftrightarrow P(S/x) > 0.5$$

ANN & a priori $\mathcal{P}$

Evis < 200 ADC counts



Ann probability for Numu CC



Ann probability for NC

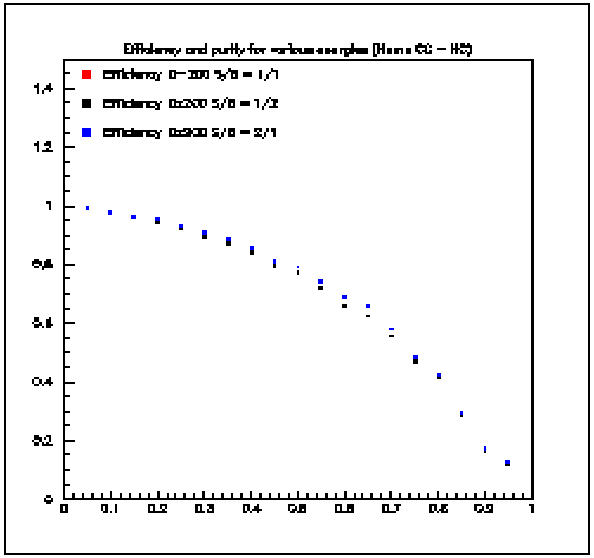
- When the S/B increases the Signal ANN probability shifts to higher values and the Background ANN probability to higher as well, since for the first it becomes most probable to be selected and for the second not.
- That means that the efficiency becomes higher and the purity lower…
- Cut @ 0.5
  - S/B 1:1 EFFICIENCY = 80 % Purity = 77 %
  - S/B 2:1 EFFICIENCY = 89 % Purity = 66 %
  - S/B 3:1 EFFICIENCY = 93 % Purity = 62 %

ANN & a priori $\mathscr{P}$

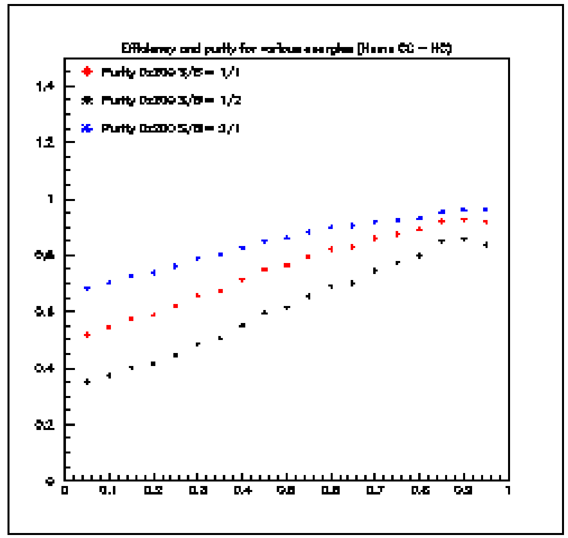
Evis < 200 ADC counts

S/B=1:1
S/B=2:1
S/B=1:2

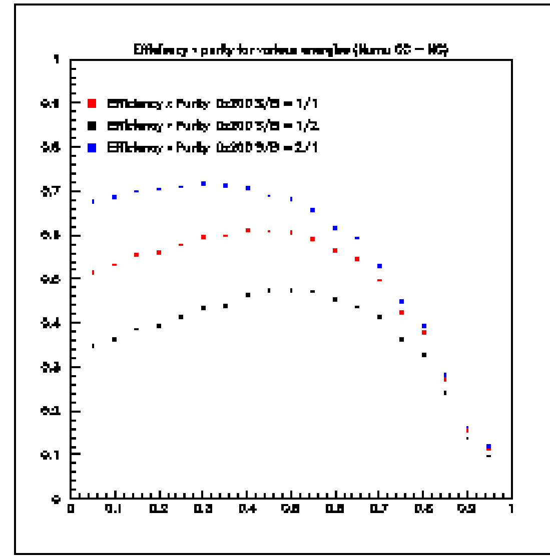
Efficiency



Purity



Eff x Pur



- When we train an ANN with a specific S/B ratio (a priori probabilities) then the efficiency we calculate does not change when we apply this ANN to any other sample of the same populations (even if the individual events probabilities we estimate are wrong).

- But the Purity depends (and thus changes) on the S/B ratio with which we train.

Summary / Comments

- When # planes > 35 only Numu CC events survive.

- Numu CC – NC separation for the PH<200, 200<PH<400 & PH>400 using ANNs is promising, given the fact that the various variables characterizing the events are highly overlapping.

- A priori probabilities are important for any kind of Bayesian (including ANNs) in order to optimize event classification and correctly calculate Signal selection efficiency and purity.

Future Work

- Redo the analysis with ~ correct a priori probabilities, and also vary those to study this effect to the classification results.
- When hadron (and electron) shower energy available use that (along with the tracks momentum) to get a better estimate of the E visible.
- Continue the similar analysis for NC – Nue CC event classification.
- Study the interpretation of ANN probabilities to CL.
- Use the results for the calculation of oscillation parameters.